

PATHOGEN TRANSMISSION

Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a regional outbreak

Evan S. Snitkin,^{1,2*} Sarah Won,³ Ali Pirani,^{1,2} Zena Lapp,⁴ Robert A. Weinstein,^{3,5} Karen Lolans,⁶ Mary K. Hayden^{3,6*}

Development of effective strategies to limit the proliferation of multidrug-resistant organisms requires a thorough understanding of how such organisms spread among health care facilities. We sought to uncover the chains of transmission underlying a 2008 U.S. regional outbreak of carbapenem-resistant *Klebsiella pneumoniae* by performing an integrated analysis of genomic and interfacility patient-transfer data. Genomic analysis yielded a high-resolution transmission network that assigned directionality to regional transmission events and discriminated between intra- and interfacility transmission when epidemiologic data were ambiguous or misleading. Examining the genomic transmission network in the context of interfacility patient transfers (patient-sharing networks) supported the role of patient transfers in driving the outbreak, with genomic analysis revealing that a small subset of patient-transfer events was sufficient to explain regional spread. Further integration of the genomic and patient-sharing networks identified one nursing home as an important bridge facility early in the outbreak—a role that was not apparent from analysis of genomic or patient-transfer data alone. Last, we found that when simulating a real-time regional outbreak, our methodology was able to accurately infer the facility at which patients acquired their infections. This approach has the potential to identify facilities with high rates of intra- or interfacility transmission, data that will be useful for triggering targeted interventions to prevent further spread of multidrug-resistant organisms.

INTRODUCTION

Infections due to multidrug-resistant organisms are a major source of health care-associated morbidity, mortality, and cost (1, 2). Because the rate of evolution of antibiotic resistance by pathogens continues to outpace the development of new antimicrobial agents, prevention of transmission of multidrug-resistant organisms is critical to the mitigation of antibiotic resistance threats (3). However, the development of effective prevention strategies is hindered by significant gaps in our understanding of the epidemiology of resistance transmission. In particular, although a great deal of effort has gone into elucidating optimal strategies to prevent transmission within hospitals, far less attention has been given to controlling dissemination of multidrug-resistant organisms across health care networks.

Patient sharing, which is the transfer of patients among different health care facilities, is increasingly recognized as an important mechanism by which multidrug-resistant organisms are spread across regions (4–7). In addition to modeling work that has demonstrated the potential for patient sharing to lead to regional dissemination of health care-associated pathogens (8), several studies have used empirical data to show that transfers of patients between health care facilities affect circulating pathogen populations as well as colonization and infection rates (9–12). These reports and others have prompted coordinated regional interventions to control multidrug-resistant organisms across health care facilities (13), with some early efforts having demonstrated impressive results (14, 15).

Although the aforementioned studies are consistent with patient-sharing networks mediating regional spread, there are few examples where the spread of multidrug-resistant organisms between facilities has been traced to the movement of specific patients (16). In one such study, we conducted an epidemiologic investigation of an outbreak of carbapenem-resistant *Klebsiella pneumoniae* (CRKP) affecting 40 patients and 26 health care facilities in four adjacent counties in Indiana and Illinois (17). Our early recognition of the outbreak and the fact that it represented the first introduction of CRKP into the region allowed us to perform a detailed review of the medical record for each case patient and assess whether patient movement between health care facilities was sufficient to explain regional spread. Strikingly, we found that virtually all case patients were linked by a patient-sharing network that connected regional acute care hospitals (ACHs), nursing homes (NHs), and long-term ACHs (LTACHs), leading us to conclude that patient sharing facilitated dissemination of CRKP across the four-county region.

The observed role of patient sharing in driving regional outbreaks raises the possibility that the analysis of patient-sharing networks can be used to guide regional infection prevention (11, 18). However, despite providing insight into potential pathways for regional spread, there are barriers to translating insights from patient-sharing networks into targeted interventions. First, although analyses of patient-sharing networks may highlight routes of interfacility transmission, they leave hidden the relative roles of intra- and interfacility transmission on the infection burden at individual facilities, which is essential knowledge for tailoring appropriate interventions. Second, it remains unclear whether an optimal strategy for regional intervention is to target health care facilities based solely on their connectivity to other facilities, or whether other facility characteristics, such as the acuity of patient populations or local infection control practices, influence a facility's regional importance. Last, given the largely static nature of patient-sharing networks, it is unclear how useful they will be in guiding interventions targeting emerging threats, or in monitoring the impact of subsequent interventions.

¹Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA. ²Division of Infectious Diseases, Department of Medicine, University of Michigan, Ann Arbor, MI 48109, USA. ³Division of Infectious Diseases, Department of Internal Medicine, Rush University Medical Center, Chicago, IL 60612, USA. ⁴Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109, USA. ⁵Division of Infectious Diseases, Department of Medicine, Cook County Health and Hospitals System, Chicago, IL 60612, USA. ⁶Division of Clinical Microbiology, Department of Pathology, Rush University Medical Center, Chicago, IL 60612, USA. *Corresponding author. Email: esnitkin@umich.edu (E.S.S.); mhayden@rush.edu (M.K.H.)

Here, we explore the use of whole-genome sequencing to track the regional spread of multidrug-resistant organisms and guide regional interventions. Sequencing isolates from the aforementioned regional outbreak of CRKP (17), we achieved sufficient resolution to map interfacility transmission pathways and to distinguish between CRKP acquisitions stemming from intra- and interfacility transmission events. Examination of the genomic transmission network in the context of patient-sharing networks confirmed the role of patient sharing in the outbreak and demonstrated how a handful of patient transfers could seed a regional outbreak.

RESULTS

Genomic analysis reveals a single importation of CRKP into the region, followed by multiple transmissions into different health care facilities

In our previous epidemiologic investigation, we applied pulsed-field gel electrophoresis, multilocus sequence typing, and patient-sharing network analysis to reach the conclusion that the regional outbreak of CRKP stemmed from a single introduction of sequence type 258 CRKP into the Chicago region of the United States in late 2007, followed by interfacility dissemination via patient-transfer events (17). We began the current genomic epidemiologic investigation by first testing the hypothesis that the outbreak did stem from one introduction of CRKP into the region. The dated phylogenetic reconstruction in Fig. 1 demonstrates that all outbreak isolates could be traced back to a common ancestor dating to mid-2007. The phylogeny also indicated early branching events (CRKP clades A, B, and C in Fig. 1), which could be consistent with a single clonal introduction and subsequent diversification in the region or with an alternative scenario of multiple contemporaneous regional

introductions of related strains. In an attempt to differentiate these two possibilities, we sequenced an additional clinical isolate that was not part of our original investigation and that represented the first CRKP-positive patient identified in the region in December 2007 (fig. S1). Mapping this isolate to the phylogeny placed it very close to the root, providing support for our hypothesis that there was a single introduction of CRKP into the region (fig. S2).

Genomic analysis confirmed the major role played by LTACH-A, but revealed additional distinct transmission pathways in the regional outbreak of CRKP

A second major conclusion from our original outbreak investigation was the central role played by LTACH-A (17), which we inferred from its early involvement and central placement in the patient-sharing network (Fig. 2). However, examining the phylogeny in Fig. 1 indicated that only one of the three major clades (clade A) was dominated by isolates from LTACH-A. Clade B was dominated by isolates from ACH-I, whose early involvement was not appreciated in the original outbreak investigation. Clade C comprised a mix of isolates from several facilities, including facilities K and H, which are disconnected from the rest of the patient-sharing network (Fig. 2), and facilities O and I, which are not proximate to each other in the patient-sharing network. Thus, it appears that clade C contained isolates that traversed the region via patient movements that were not captured in our original investigation.

Genomic analysis discriminates intra- and interfacility transmission events

The phylogeny in Fig. 1 also supports the hypothesis that the outbreak strain of CRKP was imported into each facility multiple times, with varying degrees of subsequent intrafacility transmission. The five patients from ACH-O exemplified this underlying epidemiologic complexity (patients 31, 34, 35, 38, and 39 in Fig. 1 and fig. S1). These five patients' isolates were derived from clinical cultures that were collected over a 1-month period and that were obtained >3 days after each patient's admission to ACH-O. In the absence of the whole-genome phylogeny, this cluster of infections would likely be assumed to have stemmed from a single introduction of CRKP into ACH-O with subsequent intrafacility spread. However, the placement of these isolates into three distinct phylogenetic clusters provided strong evidence that these five cases were derived from three independent introductions into the hospital, with two subsequent intrafacility transmission events (patients 35/39 and 34/38) (Fig. 1).

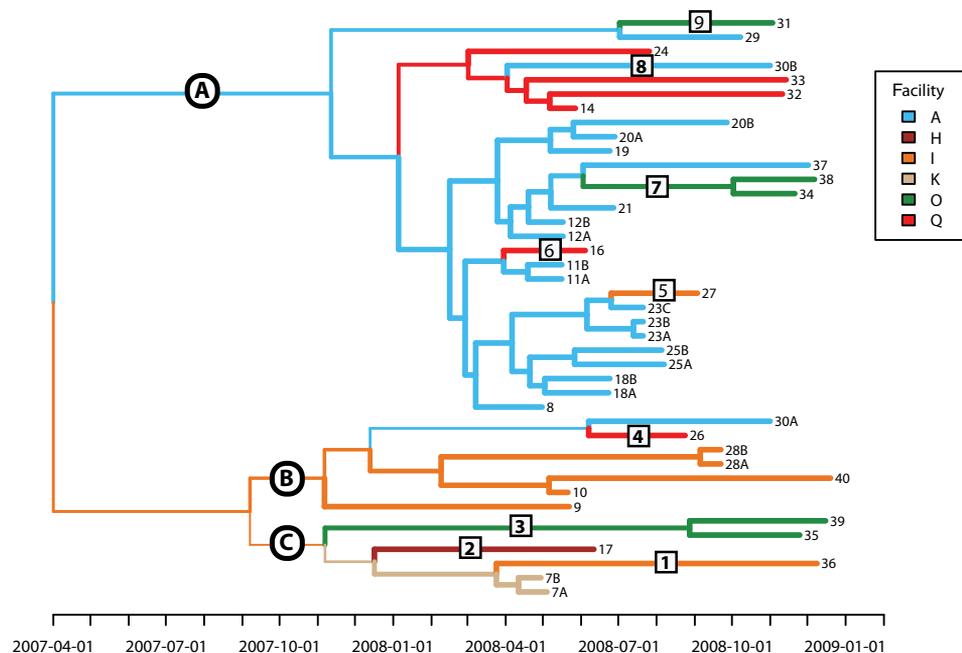


Fig. 1. Ancestral reconstruction of the 2008 regional carbapenem-resistant *K. pneumoniae* outbreak. Phylogenetic and ancestral facility reconstruction was performed in BEAST using variants, indels, health care facility locations, and dates of isolation for 41 isolates from 31 patients. Colors of branches correspond to the most probable ancestral facility, with the branch thickness indicating the relative probability as compared to all other considered facilities. Numbers on branches correspond to nine predicted interfacility transmission events based on the intermixing of isolates from different facilities on the tree. Branches labeled A, B, and C highlight the early branching of outbreak isolates into subclades that have varying concordance with the patient-sharing network.

Specific patient-transfer events support genomically inferred interfacility transmissions

With the phylogenetic reconstruction in Fig. 1 demonstrating that whole-genome sequencing provided sufficient resolution to distinguish intra- and interfacility

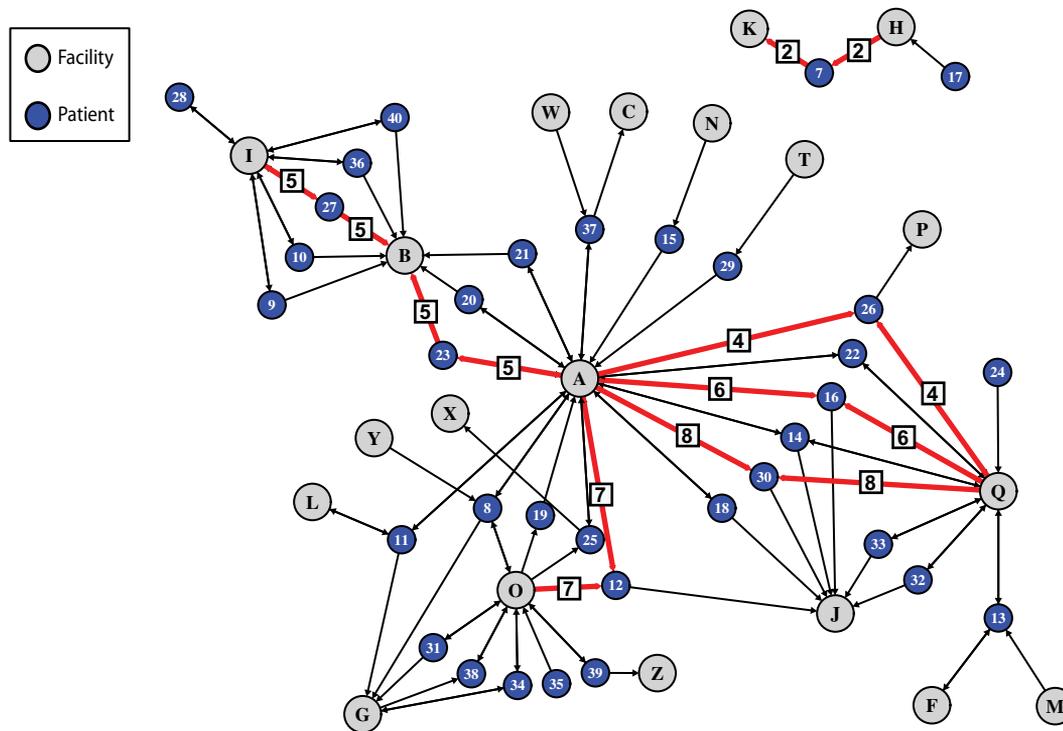


Fig. 2. A subset of patient transfers explains the genomic transmission network. The patient-transfer network from (17) was reconstructed. Blue circles represent patients, gray nodes that contain letters represent facilities, and directed edges represent patients coming from or going to a given facility. Patient-transfer events consistent with the genomic transmission network (red edges) were identified by searching for common facility exposures among patients in the subclades flanking each interfacility transmission event on the tree (see Materials and Methods and table S1). Numbers on the edges correspond to the branch labels from Fig. 1 and are only shown for the six genomic linkages that can be explained by specific patient-transfer events. Note that although a single patient-sharing explanation is displayed, in many cases, more than one transfer event could explain a genetic linkage, because patients were often transferred among common sets of facilities (for example, an acute care hospital, a nursing home, and a long-term acute care hospital) as their clinical care needs changed.

transmission events, we next set out to test the hypothesis that interfacility transmissions were due to sharing of colonized or infected patients between facilities. To this end, we extracted interfacility transmissions based on intermixing of patients' isolates from different facilities on the phylogenetic tree (Fig. 1) and asked whether patients in these mixed facility subclades shared common facility exposures (Fig. 2). In six of the nine instances (67%), patients from different facilities whose isolates clustered on the tree spent time in a common facility where cross-transmission could have occurred (Fig. 2 and table S1). Moreover, by disregarding clade C, we saw an even stronger association, with five of six predictions (83%) from clades A and B having patient-sharing explanations.

Identification of epidemiologic explanations that rely on patient sharing for ~70% of genomically inferred interfacility transmissions of CRKP would appear to support the hypothesis that patient sharing drove this regional outbreak. However, it is possible that the apparent concordance between the two data sets is simply due to the high density of connections in the patient-sharing network resulting in our identifying epidemiologically consistent explanations purely by chance. To investigate this possibility, we devised a permutation test whereby we tested whether patients from different facilities who were grouped on the phylogeny tended to overlap in the patient-sharing network

more than would be expected by chance. The results of this permutation analysis strongly support the assertion that concordance between the genomic and patient-sharing networks is nonrandom ($P = 0.001$; fig. S3).

An NH was identified as a key bridge facility early in the CRKP regional outbreak

With genomic support for the role of patient movement in the propagation of this regional outbreak of CRKP, we next sought to analyze the genomic transmission network to gain a more complete understanding of the key transmission events that mediated regional spread. Consistent with our original investigation, we observed strong support for LTACH-A playing a central role, with five of nine interfacility transmissions originating from LTACH-A (table S1). However, our genomic transmission network also supported a previously unappreciated role for ACH-I early in the outbreak, with a cluster of isolates from four patients in ACH-I branching away from the clade containing all LTACH-A isolates (Fig. 1 and fig. S2). To understand whether and how the early

branching clusters in ACH-I and LTACH-A might be connected, we re-examined the patient-sharing network to search for links between these two facilities. This inspection revealed no direct transfers of patients between LTACH-A and ACH-I. However, both LTACH-A and ACH-I had many connections with a third facility, NH-B (Fig. 2), raising the possibility that NH-B may have played a role in the outbreak as the index facility or as an intermediary between LTACH-A and ACH-I. To test this hypothesis, we sequenced an additional set of CRKP surveillance isolates from 6 patients from NH-B and 15 patients from LTACH-A, who were not considered in our previous investigation. Supporting the connection between ACH-I and NH-B, five of the six NH-B isolates were intermixed in the phylogeny with the ACH-I cluster, with the sixth isolate falling into the LTACH-A subclade (fig. S4). Updating the genomic transmission network with these isolates supported the hypothesis that NH-B played a central role early in the outbreak, potentially seeding transmission clusters in both ACH-I and LTACH-A (fig. S4).

Frequency of patient transfers between facilities predicts frequency of genomically inferred CRKP transmissions

With what we now perceived to be a complete picture of the interfacility transmissions that seeded the outbreak, we next asked what determined

the frequency of CRKP transmissions between regional facilities. In particular, we wondered whether the frequency of transmission between facilities could be predicted by the number of patients moving between those facilities, independent of knowledge of other factors such as characteristics of a facility's patient population or its infection control practices. To investigate this question, we compared the structure and density of links in the genomically inferred CRKP transmission network and the patient-sharing network (Fig. 3A). We observed that both the structure and density of these networks were highly correlated ($P = 0.027$, Mantel test; Fig. 3B), supporting the hypothesis that the density of patient transfers alone was predictive of interfacility CRKP transmission frequency.

Real-time analysis yields accurate insights into the origin of patients' CRKP isolates

Our results demonstrate that whole-genome sequencing and phylogenetic analysis can be applied to infer both intra- and interfacility trans-

missions of CRKP during a regional outbreak (Fig. 3). Although this points to the utility of such an approach for retrospective studies of outbreaks, we next asked whether the analytical approach that we took here could be applied to study an outbreak in real time. To test this hypothesis, we evaluated how well we could predict the facility of origin for each patient's CRKP isolate when using only isolates from other patients that preceded that patient in the outbreak. For all cases of CRKP, we found that the prediction using only real-time data matched the prediction using the entire data set (Fig. 4).

DISCUSSION

Here, we report on the integrated analysis of genomic and patient-sharing networks for a regional outbreak of CRKP that affected 26 health care facilities in the Midwest United States over the course of a year. We found that our genomic analysis yielded a high-resolution transmission map that allowed for discrimination between intra- and interfacility transmission events, provided insight into the direction of transmission between facilities, and verified the estimated date of introduction of CRKP to the region. Comparison of genomic and patient-sharing networks provided strong support for patient sharing having mediated regional spread, with most interfacility genomic linkages explainable by specific patient-transfer events. Whole-genome sequencing further allowed confirmation of key inferences from our original outbreak investigation (17), for example, the importance of LTACH-A in amplifying and disseminating CRKP across the region. More importantly, whole-genome sequencing allowed us to extend our original findings by identifying previously unappreciated complexity, such as multiple introductions of CRKP into ACH-O, and the early role of NH-B in propagating the outbreak. Last, real-time analysis yielded insights identical to those of the full, retrospective study, highlighting the promise of whole-genome sequencing to accelerate the pace of outbreak containment interventions.

Our findings have practical implications for control of multidrug-resistant organisms such as CRKP. From the standpoint of the individual health care facility, the ability of combined genomic and epidemiologic analysis to distinguish between intrafacility spread and repeated importation could facilitate implementation of the most effective infection prevention measures specific to each situation. From a regional public-health standpoint, emerging pathogen surveillance that incorporates whole-genome sequencing could help to identify facilities within a network that might be contributing disproportionately to multidrug-resistant organism spread, thus allowing targeted investigation and intervention. As shown by the example of NH-B in the present analysis, an integrated examination of genomic and patient-sharing networks can lead to actionable hypotheses that are not apparent by examining either data set alone.

Whereas our analysis points to the potential of genome sequencing to aid in regional control of health care-associated pathogens, there are caveats that should be considered. First, although we studied a multidrug-resistant bacterium that is endemic in many areas, we focused on its dissemination after initial introduction into one region of the United States. It will be important in future studies to evaluate how effective whole-genome sequencing is in quantifying intra- and interfacility transmission links in the more common endemic situation when a multidrug-resistant organism has had the opportunity to move between facilities many times. Although this will be more analytically challenging and will likely provide a more complex genomic and epidemiologic picture, previous investigations have shown impressive

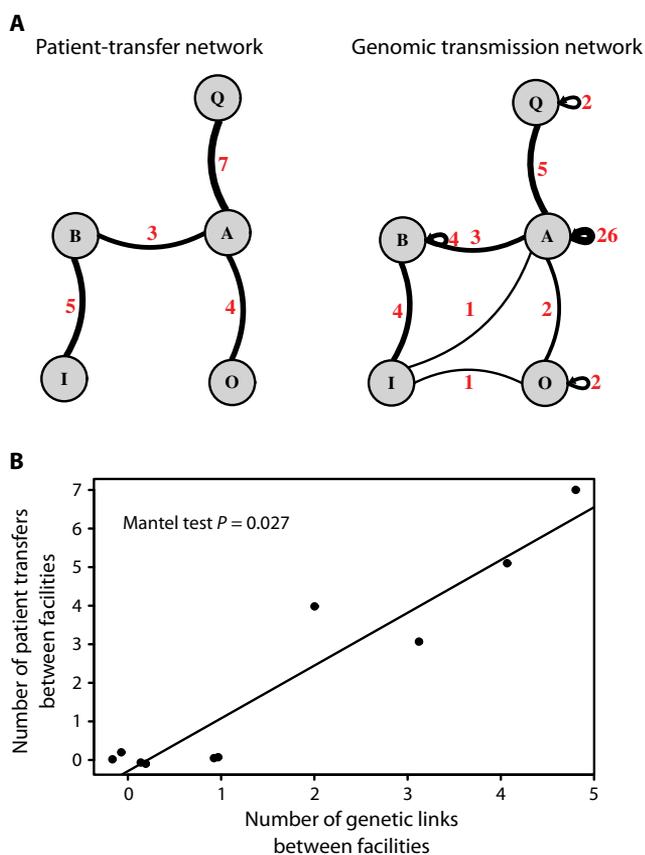


Fig. 3. Comparison between genomic and patient-transfer networks. (A) The number of interfacility linkages observed in the patient-transfer network (left) and genomic transmission network (right) is shown as undirected networks. For the genomic network, the number of predicted intrafacility transmissions is also indicated on loops circling back to each facility label. Genomic transmissions were extracted from the phylogenetic reconstruction that included the set of surveillance isolates from NH-B (fig. S4). Note that the patient-sharing network does not consider the additional set of surveillance isolates from patients from NH-B and LTACH-A, because the patients' history of health care exposures was not known. (B) The number of connections between each pair of facilities from the patient transfer and genomic networks in (A) is compared in a scatterplot, with each point representing a pair of facilities. Points on the plot are jittered, such that overlapping points can be distinguished. The number of interfacility connections in the two networks was compared using a Mantel test ($P = 0.027$).

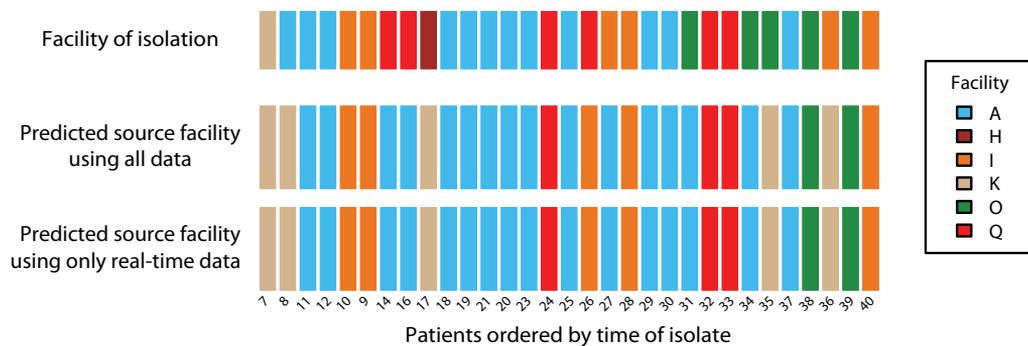


Fig. 4. Comparison of real-time and retrospective genomic predictions of patients' isolate origins. Genomic predictions for the origin of each patient's isolate were compared by using the complete data set and real-time subsets of the data. Real-time data subsets consisted of only isolates that predated a patient's isolate for which a prediction was being made. Patients are ordered from left to right by the date on which the specimen from which their isolate was grown was collected. The top color bar indicates the facility where the patient resided when the culture from which their carbapenem-resistant *K. pneumoniae* isolate was collected, the middle color bar indicates the predicted source facility by using only real-time data, and the bottom color bar indicates the predicted source facility when all data were considered. Differences from the top color bar indicate a predicted importation, and conversely, identities indicate a predicted intrafacility transmission. Note that the prediction for the first isolate is trivial, because there was no preceding isolate to link it to.

precision in tracking regional spread of pathogens over time (19). Second, whereas CRKP in the United States is thought to be exclusively health care-associated, other antibiotic-resistant bacteria may have transmission pathways that extend into the community, potentially limiting the utility of patient-transfer networks in interpreting regional transmission events. However, we believe that the broad approach of integrating genetic and epidemiologic data to discern regional transmission pathways will be generalizable to multidrug-resistant organisms with different underlying epidemiology but will require high-quality epidemiologic data that probe the transmission pathways relevant to the organism under study. For example, we recently found that whereas health care exposures did not explain the genetic clustering of methicillin-resistant *Staphylococcus aureus* strain USA300 bloodstream infection isolates, community factors (for example, race and ethnicity) did, which indicates that community transmission networks may be a primary driver of USA300 strain proliferation (20). Third, our study benefited from having a relatively complete set of isolates from affected facilities. Constructing an accurate regional transmission network is dependent on having representative sampling, which may not always be possible, especially with emerging microbial threats. However, we also demonstrated how combining genomic analysis with robust epidemiologic data can mitigate the impact of incomplete sampling as well as facilitate inferences into the putative importance of facilities from which isolates are unavailable (for example, NH-B in fig. S4) and identify facilities for which additional epidemiologic investigation is required (for example, clade C in Fig. 1). Finally, although accurate interfacility transmission predictions can be made when simulating real-time availability of isolates, there are additional hurdles that will need to be surmounted. Sequencing and bioinformatics obstacles are rapidly being overcome, but effective surveillance of health care systems will require participation from all parts of the health care network, timely reporting, and an infrastructure capable of identifying and responding to dynamic threats in real time.

In summary, through the integration of high-resolution molecular and epidemiologic data sets, we have built on previous large-scale analyses of health care networks (9–12) and demonstrate how individual transfer events can lead to regional spread of an emerging multidrug-

resistant organism (in this case, CRKP). We expect that future genomic-epidemiologic investigations encompassing larger health care networks and considering endemic pathogens will further improve our understanding of how multidrug-resistant organisms permeate health care systems, which will help to guide interventions to eradicate them.

MATERIALS AND METHODS

Study design

This study was a retrospective genomic epidemiologic investigation of a regional outbreak of CRKP. The objectives of the study were (i) to evaluate the capacity of whole-genome sequencing to delineate interfacility transmission pathways for multidrug-resistant

organisms, (ii) to compare genomic and patient-sharing networks to evaluate the role of patient movement in mediating the regional spread of CRKP, and (iii) to evaluate the potential for genomic surveillance of high-priority multidrug-resistant organisms to lead to actionable insights.

Whole-genome sequencing

DNA was extracted with the MoBio PowerMag Microbial DNA kit and prepared for sequencing on an Illumina MiSeq instrument using the NEBNext Ultra kit and sample-specific barcoding. Library preparation and sequencing were performed at the Center for Microbial Systems at the University of Michigan. Quality of reads was assessed with FastQC (21), and Trimmomatic (22) was used for trimming adapter sequences and low-quality bases. Assemblies were performed using the A5 pipeline with default parameters (23).

Variants were identified by (i) mapping filtered reads to the assembled H-1 genome (see table S2) using the Burrows-Wheeler short-read aligner, (ii) discarding polymerase chain reaction duplicates with Picard, and (iii) calling variants with SAMtools and bcftools. Variants were filtered from raw results using GATK's VariantFiltration (QUAL, >100; MQ, >50; >10 reads supporting variant; and FQ, <0.025). In addition, a custom python script was used to filter out single-nucleotide variants that were (i) <5 base pairs (bp) in proximity to indels, (ii) <10 bp in proximity to another variant, (iii) not present in the core genome, or (iv) in a recombinant region identified by Gubbins (fig. S5) (24). Large genomic deletions were identified from mapped reads using bedtools, with 5-kb nonoverlapping windows. For purposes of phylogenetic analyses, contiguous windows showing the same pattern of presence and absence across all genomes were consolidated to single deletion events.

Phylogenetic analysis

Maximum likelihood trees were constructed with variants and indels in RAxML (25) by performing a partitioned analysis wherein variants were modeled with a general-time reversible (GTR) model and indels as uncorrelated binary characters. Bootstrap analysis was performed with the number of bootstrap replicates determined using the bootstrap

convergence test and the autoMRE convergence criteria (-N autoMRE). Bootstrap support values were overlaid on the best-scoring tree identified during rapid bootstrap analysis (-f a).

The molecular clock hypothesis was tested among outbreak genomes using Path-O-Gen by correlating root-to-tip distance in the RAxML tree and dates of isolates (fig. S6) (26). Dated phylogenetic trees were constructed using BEAST v2.4.5 (27). The data were partitioned into single-nucleotide variants and indels with a linked tree model and unlinked substitution and clock models. BEAST was run using dated tips, the GTR model of nucleotide substitution for variants, the mutation death model of substitution for indels, a strict molecular clock model for variants and indels, and a coalescent exponential population prior. The mutation death model assumes that a deletion is more likely than an insertion. To test how clock-like the variants and indels are, a lognormal relaxed clock was used with a γ distribution prior (median, 0.1) on the SD. For variants and indels, the probability mass was concentrated near zero; thus, a strict clock model cannot be rejected. The coalescent exponential population prior was chosen because the marginal posterior distribution of the growth rate did not include zero, indicating that the data are not compatible with a constant population size. Isolate facility was added as a discrete trait. Three independent BEAST runs of 50 million generations, each with a burn-in of 10%, were combined to get an effective sample size (ESS) of >900 for all parameters. The maximum clade credibility consensus tree with marginal probabilities of location at each node was generated using mean heights in TreeAnnotator.

Construction and analysis of interfacility transmission networks

Regional transmission networks were constructed by first calculating the relative likelihoods of unobserved ancestral outbreak strains being associated with each facility. Ancestral reconstruction was performed using BEAST or the maximum likelihood rerooting method, as implemented in phytools (28). The maximum likelihood approach was used in the permutation (see below and fig. S3) and real-time analyses (Fig. 4) because of its lower computational demands. Interfacility transmissions were extracted from the ancestrally reconstructed phylogeny if a branch connected two nodes with different ancestral state assignments, with both ancestral assignments having more than half of the marginal likelihood assigned to them. Intrafacility transmissions were tabulated as branches connecting tips to an internal node with the same ancestral state. To account for importation into facilities, the number of transmissions into a given facility was subtracted from the total intrafacility transmission count. Predictions of interfacility transmission linkages were highly concordant between BEAST and the maximum likelihood approach (fig. S7).

For comparison to the patient-sharing network, only interfacility links involving subclades with isolates that were uniform for the facility of isolation were considered. This decision was made to focus on interfacility transmissions that could be explained by a single patient-sharing event. To identify patient-sharing explanations for interfacility linkages, we designated source and destination subclades for each interfacility transmission extracted from the tree. Next, we examined the patient-sharing network for any of the following three types of epidemiologic linkages: (i) a patient in the destination subclade having previously spent time in the source facility, (ii) a patient in the source subclade having previously spent time in the destination facility, or (iii) a patient in the destination subclade and a patient in the source subclade having both spent time in some other common facility. To evaluate whether

observed concordance between genetic and patient-sharing networks was nonrandom, we devised a permutation test, wherein for each iteration, each patient in the tree was assigned another random patient's transfer history. For each of 1000 permutations, we (i) performed a new ancestral reconstruction based on each patient's new facility assignment, (ii) identified interfacility transmission links from the ancestrally reconstructed phylogeny, and (iii) determined the number of interfacility linkages with patient-sharing explanations.

Statistical analysis

Applied statistical tests included Pearson correlation analysis for testing the molecular clock hypothesis, Mantel tests for comparing the structure of patient-sharing and genomic networks, and custom permutation tests for evaluating the capacity of patient-sharing events to explain interfacility genomic linkages. All statistical analyses were performed in R.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/9/417/eaan0093/DC1

Fig. S1. Epidemic curve of outbreak across regional health care facilities.

Fig. S2. Maximum likelihood tree of outbreak isolates.

Fig. S3. Distribution of overlaps between genomic and patient-sharing networks for randomized data.

Fig. S4. Ancestral facility reconstruction for collection including surveillance isolates from NH-B and LTACH-A.

Fig. S5. Variants identified among outbreak isolates.

Fig. S6. Correlation between date of isolation and genetic distance from common ancestor.

Fig. S7. Maximum likelihood ancestral reconstruction of outbreak isolates.

Table S1. Patient-transfer events associated with genomically inferred interfacility transmissions.

Table S2. Summary of whole-genome sequencing in the current study.

REFERENCES AND NOTES

- G. M. Eliopoulos, L. L. Maragakis, T. M. Perl, *Acinetobacter baumannii*: Epidemiology, antimicrobial resistance, and treatment options. *Clin. Infect. Dis.* **46**, 1254–1263 (2008).
- A. Huttner, S. Harbarth, J. Carlet, S. Cosgrove, H. Goossens, A. Holmes, V. Jarlier, A. Voss, D. Pittet; World Healthcare-Associated Infections Forum participants, Antimicrobial resistance: A global view from the 2013 World Healthcare-Associated Infections Forum. *Antimicrob. Resist. Infect. Control* **2**, 31 (2013).
- H. W. Boucher, G. H. Talbot, J. S. Bradley, J. E. Edwards, D. Gilbert, L. B. Rice, M. Scheld, B. Spellberg, J. Bartlett, Bad bugs, no drugs: No ESCAPE! An update from the Infectious Diseases Society of America. *Clin. Infect. Dis.* **48**, 1–12 (2009).
- M. Ciccolini, T. Donker, R. Köck, M. Mielke, R. Hendrix, A. Jurke, J. Rahamat-Langendoen, K. Becker, H. G. M. Niesters, H. Grundmann, A. W. Friedrich, Infection prevention in a connected world: The case for a regional approach. *Int. J. Med. Microbiol.* **303**, 380–387 (2013).
- R. A. Weinstein, S. A. Kabins, Strategies for prevention and control of multiple drug-resistant nosocomial infection. *Am. J. Med.* **70**, 449–454 (1981).
- J. Wiener, J. P. Quinn, P. A. Bradford, R. V. Goering, C. Nathan, K. Bush, R. A. Weinstein, Multiple antibiotic-resistant *Klebsiella* and *Escherichia coli* in nursing homes. *JAMA* **281**, 517–523 (1999).
- W. E. Trick, M. J. Kuehnert, S. B. Quirk, M. J. Arduino, S. M. Aguero, L. A. Carson, B. C. Hill, S. N. Banerjee, W. R. Jarvis, Regional dissemination of vancomycin-resistant enterococci resulting from interfacility transfer of colonized patients. *J. Infect. Dis.* **180**, 391–396 (1999).
- T. Donker, J. Wallinga, H. Grundmann, Patient referral patterns and the spread of hospital-acquired infections through national health care networks. *PLOS Comput. Biol.* **6**, e1000715 (2010).
- W. Ke, S. S. Huang, L. O. Hudson, K. R. Elkins, C. C. Nguyen, B. G. Spratt, C. R. Murphy, T. R. Avery, M. Lipsitch, Patient sharing and population genetic structure of methicillin-resistant *Staphylococcus aureus*. *Proc. Natl. Acad. Sci.* **109**, 6763–6768 (2012).
- T. Donker, J. Wallinga, R. Slack, H. Grundmann, Hospital networks and the dispersal of hospital-acquired pathogens by patient transfer. *PLOS ONE* **7**, e35002 (2012).
- M. J. Ray, M. Y. Lin, R. A. Weinstein, W. E. Trick, Spread of carbapenem-resistant *Enterobacteriaceae* among Illinois healthcare facilities: The role of patient sharing. *Clin. Infect. Dis.* **63**, 889–893 (2016).
- H.-H. Chang, J. Dordel, T. Donker, C. J. Worby, E. J. Feil, W. P. Hanage, S. D. Bentley, S. S. Huang, M. Lipsitch, Identifying the effect of patient sharing on between-hospital

- genetic differentiation of methicillin-resistant *Staphylococcus aureus*. *Genome Med.* **8**, 18 (2016).
13. R. B. Slayton, D. Toth, B. Y. Lee, W. Tanner, S. M. Bartsch, K. Khader, K. Wong, K. Brown, J. A. McKinnell, W. Ray, L. G. Miller, M. Rubin, D. S. Kim, F. Adler, C. Cao, L. Avery, N. T. B. Stone, A. Kallen, M. Samore, S. S. Huang, S. Fridkin, J. A. Jernigan, Vital signs: Estimated effects of a coordinated approach for action to reduce antibiotic-resistant infections in health care facilities—United States. *MMWR Morb. Mortal. Wkly Rep.* **64**, 826–831 (2015).
 14. D. Ben-David, S. Masarwa, A. Adler, H. Mishali, Y. Carmeli, M. J. Schwaber; Post-Acute Care Hospital Carbapenem-Resistant Enterobacteriaceae Working Group, A national intervention to prevent the spread of carbapenem-resistant Enterobacteriaceae in Israeli post-acute care hospitals. *Infect. Control Hosp. Epidemiol.* **35**, 802–809 (2014).
 15. M. J. Schwaber, Y. Carmeli, An ongoing national intervention to contain the spread of carbapenem-resistant Enterobacteriaceae. *Clin. Infect. Dis.* **58**, 697–703 (2014).
 16. K. Zhou, M. Lokate, R. H. Deurenberg, M. Tepper, J. P. Arends, E. G. C. Raangs, J. Lo-Ten-Foe, H. Grundmann, J. W. A. Rossen, A. W. Friedrich, Use of whole-genome sequencing to trace, control and characterize the regional expansion of extended-spectrum β -lactamase producing ST15 *Klebsiella pneumoniae*. *Sci. Rep.* **6**, 20840 (2016).
 17. S. Y. Won, L. S. Munoz-Price, K. Lolans, B. Hota, R. A. Weinstein, M. K. Hayden; Centers for Disease Control and Prevention Epicenter Program, Emergence and rapid regional spread of *Klebsiella pneumoniae* carbapenemase-producing *Enterobacteriaceae*. *Clin. Infect. Dis.* **53**, 532–540 (2011).
 18. M. Ciccolini, T. Donker, H. Grundmann, M. J. M. Bonten, M. E. J. Woolhouse, Efficient surveillance for healthcare-associated infections spreading between hospitals. *Proc. Natl. Acad. Sci.* **111**, 2271–2276 (2014).
 19. M. T. G. Holden, L.-Y. Hsu, K. Kurt, L. A. Weinert, A. E. Mather, S. R. Harris, B. Strommenger, F. Layer, W. Witte, H. de Lencastre, R. Skov, H. Westh, H. Žemličková, G. Coombs, A. M. Kearns, R. L. R. Hill, J. Edgeworth, I. Gould, V. Gant, J. Cooke, G. F. Edwards, P. R. McAdam, K. E. Templeton, A. McCann, Z. Zhou, S. Castillo-Ramirez, E. J. Feil, L. O. Hudson, M. C. Enright, F. Balloux, D. M. Aanensen, B. G. Spratt, J. R. Fitzgerald, J. Parkhill, M. Achtman, S. D. Bentley, U. Nübel, A genomic portrait of the emergence, evolution and global spread of a methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res.* **23**, 653–664 (2013).
 20. K. J. Popovich, E. S. Snitkin, B. Hota, S. J. Green, A. Pirani, A. Aroutcheva, R. A. Weinstein, Genomic and epidemiological evidence for community origins of hospital-onset methicillin-resistant *Staphylococcus aureus* bloodstream infections. *J. Infect. Dis.* **215**, 1640–1647 (2017).
 21. S. Andrews, *FastQC: A quality control tool for high throughput sequence data*; www.bioinformatics.babraham.ac.uk/projects/fastqc.
 22. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
 23. D. Coil, G. Jospin, A. E. Darling, A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* **31**, 587–589 (2015).
 24. N. J. Croucher, A. J. Page, T. R. Connor, A. J. Delaney, J. A. Keane, S. D. Bentley, J. Parkhill, S. R. Harris, Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res.* **43**, e15 (2015).
 25. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
 26. A. Rambaut, T. T. Lam, L. Max Carvalho, O. G. Pybus, Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
 27. R. Bouckaert, J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu, D. Xie, M. A. Suchard, A. Rambaut, A. J. Drummond, BEAST 2: A software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **10**, e1003537 (2014).
 28. L. J. Revell, phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
- Acknowledgments:** We would like to acknowledge S. Whitefield for assistance with sample preparation and helpful comments on analytical approaches. **Funding:** This work was supported by the Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program (U54 CK000161 0452). Z.L. received support from training grant T32 HG00040 and the NSF Graduate Research Fellowship Program under grant no. DGE 1256260. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF. **Author contributions:** E.S.S., M.K.H., and S.W. designed the study, did the literature search, and wrote the manuscript. E.S.S., S.W., K.L., and M.K.H. conducted the study and collected the data. E.S.S. and A.P. processed and analyzed raw genomic data. E.S.S. and Z.L. performed genomic epidemiology analyses. E.S.S., S.W., M.K.H., and R.A.W. interpreted the results. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Sequence data are available under BioProject #PRJNA355241 (www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA355241).
- Submitted 17 February 2017
Resubmitted 16 June 2017
Accepted 18 August 2017
Published 22 November 2017
10.1126/scitranslmed.aan0093
- Citation:** E. S. Snitkin, S. Won, A. Pirani, Z. Lapp, R. A. Weinstein, K. Lolans, M. K. Hayden, Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a regional outbreak. *Sci. Transl. Med.* **9**, eaan0093 (2017).

Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a regional outbreak

Evan S. Snitkin, Sarah Won, Ali Pirani, Zena Lapp, Robert A. Weinstein, Karen Lolans and Mary K. Hayden

Sci Transl Med **9**, eaan0093.
DOI: 10.1126/scitranslmed.aan0093

Drug-resistant out-of-towners

With the rise of multidrug-resistant organisms, much attention has focused on preventing nosocomial infection within individual health care facilities. Transfer of patients between facilities, however, may also contribute to the proliferation of hard-to-treat bacterial pathogens. Snitkin *et al.* investigated the role of patient transfer between hospitals in a year-long regional outbreak of carbapenem-resistant *Klebsiella pneumoniae* in the United States. They traced the spread of the infection by combining genomic sequencing with information about the movement of patients across the local health care network. This analysis was able to identify where the outbreak began and pinpointed which facilities contributed to the transmission of the infection in the region. Such an approach may enable directed interventions to prevent the transfer of drug-resistant organisms among hospitals.

ARTICLE TOOLS	http://stm.sciencemag.org/content/9/417/eaan0093
SUPPLEMENTARY MATERIALS	http://stm.sciencemag.org/content/suppl/2017/11/20/9.417.eaan0093.DC1
RELATED CONTENT	http://stm.sciencemag.org/content/scitransmed/10/452/eaar6115.full http://stm.sciencemag.org/content/scitransmed/11/487/eaat8418.full http://science.sciencemag.org/content/sci/365/6459/eaaw1944.full
REFERENCES	This article cites 27 articles, 3 of which you can access for free http://stm.sciencemag.org/content/9/417/eaan0093#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the [Terms of Service](#)

Science Translational Medicine (ISSN 1946-6242) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Translational Medicine* is a registered trademark of AAAS.

Copyright © 2017 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works